

GPS からの習慣モデル獲得法と「気づき」提供サービス構成法

松尾宣夫^{†1} 高見一正^{†1&†2}

近年、ブロードバンド環境の発展によるデータ量の増加、携帯端末の高機能化、SNS 等の発展により、ビッグデータやライフログが注目を集めている。ビッグデータやライフログの活用が実現すれば、有益な付加価値の提供が可能になる。しかしながら、膨大で多種多様なデータの収集、分析、管理には新たな発想と技術が求められている。一方で、ライフログの中でも位置情報、特に GPS(global positioning system)は、携帯電話やスマートフォンに搭載されており、ユーザが意識せずに比較的容易にログの取得が可能である。本稿では、携帯電話やスマートフォンでの利用を想定したライフログサービスを実現するために、GPS から得られるログ情報を解析し、有意な情報を抽出し、その有意な情報から習慣モデルの導出を行い、導出した習慣モデルとユーザの状態（時間と場所等）に基づいた「気づき」提供サービス構成法を提案する。また、本提案の有効性を示すために、既存サービスと習慣モデルを利用した「気づき」提供サービスの比較を行った。

A Habit Model and it's Capturing Method from GPS Data for Awareness Services

NOBUO MATSUO^{†1} KAZUMASA TAKAMI^{†1&†2}

Recently, big data and life log technology expected by increasing dramatically the amount of data, the progress of the mobile terminal and the spread of SNS (social network service). We are able to create the value by analyzing big data and life log. However, it is difficult to analyze and exploit a large amount of kind of the big data and the life log. On the other hand, because cell phone and smart phone are equipped with a GPS (global positioning system), it is easy to get and analyze GPS data only. So we focused on GPS in life log. In this paper, we propose configuration method for awareness services using habit model extracted from GPS. In order to confirm effectiveness our proposed solution, we compared the existing services with the awareness services of using habit model and evaluated our proposed solution.

1. はじめに

近年、ブロードバンド環境の発展によるデータ量の増加、携帯端末の高機能化、SNS 等の発展により、ビッグデータやライフログが注目を集めている。ビッグデータとは「ライフログ等の膨大で多種多様な不定形のデジタルデータ」であり、ライフログとは「個人に関する情報をデジタルデータとして記録すること、またそのデータ」である。これらのデータの活用が実現すれば、有益な付加価値を提供することが可能になる。しかし、2011年のビッグデータの総量は1兆8000億GBにも上り、膨大で多種多様なデータを連携、分析、管理するためには新たな発想と技術が求められている。さらに、ライフログにおいても人の一生を記録するには3TBの容量が必要であり、データの種類も多種多様で、記録を持続できる技術が必要である。ライフログの中でも主要なライフログとして位置情報がある。携帯電話やスマートフォンには標準でGPS(global positioning system)が搭載されており、比較的ログが取得しやすい。その上、ユーザが手動でログを記録する必要がない。そこで、研究対象のライフログをGPSに限定し、GPSログ情報の分析、管理、活用を考える。

本研究では、携帯電話やスマートフォンでの利用を想定したライフログサービスを実現するために、GPSから得られるログ情報を解析し、有意な情報を抽出し、抽出した有意な情報から習慣モデルの導出を行い、導出した習慣モデルとユーザの状態（時間と場所等）に基づいた「気づき」提供サービス構成法を提案する。

本稿では、まず第2章では関連研究について述べる。そして、第3章では本研究における習慣モデルと「気づき」提供サービスについての概要を述べる。第4章では習慣モデルを利用した「気づき」提供サービスを実現するために解決する必要がある課題について述べる。第5章では第4章で提示した課題に対する解決策について述べる。第6章では、第5章で提案した解決策の評価について述べる。最後に第7章で、まとめと今後の課題について述べる。

2. 関連研究

ライフログの研究分野には、ライフログの収集・分析、ライフログの管理、ライフログの活用がある。ライフログの収集・分析に関する研究[1][2]は数多く存在するが、ライフログの管理、ライフログの活用に関する研究[3][4]はほとんど存在しない。また、GPSログ情報の収集・分析に関する関連研究では、主にユーザの滞在地を抽出し、その滞在地の遷移関係をモデル化する研究が散見される。例えば、AshbrookらはGPSからk-meansクラスタリング手法を利用

^{†1} 創価大学大学院工学研究科情報システム工学専攻
Graduate School of Engineering, Soka University

^{†2} 創価大学工学部情報システム工学科
Faculty of Engineering Soka University

して滞在地を抽出し、抽出した滞在地の遷移関係をマルコフモデルにより確率を付与することでユーザの行動の予測を提案している[5].

また、西野らは DBSCAN クラスタリング手法を利用して滞在地を抽出し、抽出した滞在地の遷移関係を時系列順に整理したものを蓄積し、蓄積したものに頻出系列マイニングを適応することで、ユーザの頻出する滞在地の遷移関係をモデル化している[6]. どちらの研究も滞在地を特定し、特定した滞在地の遷移関係をモデル化する研究である. これらの研究では、GPSデータの分析に主眼が置かれており、分析結果をどのように管理し、活用するかという点にまで言及していない. また、遷移関係に関する情報である移動手段、移動経路等との連携も考慮されていない. 本研究と関連研究との違いは、滞在地の粒度が施設単位であること、滞在地間の遷移関係を単なるモデル化だけでなく、データ管理方法としても利用していることである. また、利用する情報も滞在地と滞在時間だけでなく、曜日等の時間データ、滞在地間の遷移情報である移動手段や移動経路を利用することも考慮している点である. そして、ログデータの収集、管理、活用というすべての分野を考慮した研究を行っている点である.

3. 習慣モデルと「気づき」提供サービス

3.1 習慣モデルと「気づき」

習慣モデルとは人の行動を滞在した場所の遷移関係で表現したものであり、日々の滞在した場所の遷移関係を蓄積して合成したモデルである. 「気づき」はユーザが次に行うであろう行動を促したり、注意を喚起したりすることである. 以下の図1に習慣モデルの概要を示す.

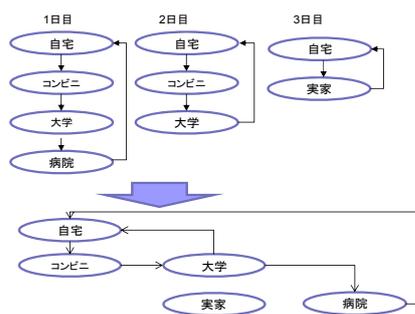


図1 習慣モデル概要
Figure 1 Outline of habit model.

3.2 習慣モデルに基づく「気づき」提供サービス構成

本研究のライフログサービスは「気づき」を与える基本サービスと「気づき」に基づいてユーザが行動を起こす時に、その行動に関連する情報提供を行う付加サービスから構成されている. 本研究で想定している「気づき」を示す.

- 終電や終バスを通知する「気づき」
- レンタル店や図書館の延滞防止を通知する「気づき」

- 過去と現在の変化を通知する「気づき」(店舗移転等)
- 生活リスク回避支援サービス

4. 課題

習慣モデルを利用した「気づき」提供サービスを実現するために解決する必要がある課題について以下に述べる.

4.1 GPS ログ情報からの有意情報の抽出

GPS ログは時刻情報や緯度・経度情報の単なる蓄積である. このようなGPS ログからどのような有意な情報を取得できるか明確にする必要がある. また、明確にした有意な情報を抽出するための手法を検討する必要がある.

4.2 習慣モデルの定義と導出

ユーザの行動を予測し、支援するために習慣モデルが必要となる. 習慣モデルを定義する上で、GPS ログから取得した情報から人の行動をどう表現することができるか検討する必要がある. また、定義した習慣モデルをどのように導出するのか考える必要がある. そして、ユーザの行動を予測するために、習慣モデル上の遷移先を決定する機能を検討する必要がある.

4.3 習慣モデルに基づくサービス提供法

導出した習慣モデルとユーザの現在の状態(時間と場所等)を利用してユーザに「気づき」サービスを提供する仕組みを検討する必要がある.

5. 解決策

5.1 GPS ログ収集実験と得られた知見

人の行動は(移動) → (施設滞在) → (移動)を繰り返している. この視点からGPS ログ収集実験を行い、収集したGPS ログ情報を分析した結果、速度と方向に関する知見を得た. 図2は本学のニューロール(レストラン)に滞在した前後のログ情報をGoogle mapとGoogle earthにマッピングした図である. 図2のように、(移動) → (施設滞在) → (移動)を意識して、GPS ログ収集実験を行った.

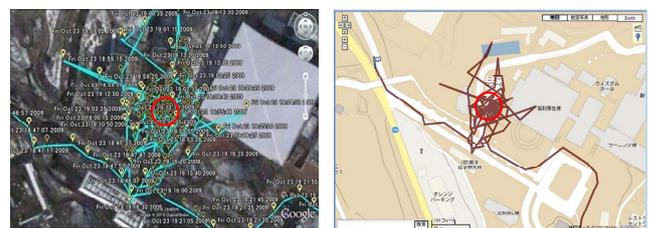


図2 GPS Loggerを使用して収集したGPS ログ情報
Figure 2 Corrected GPS log using GPS logger.

(1) 速度に関する知見

図3のグラフは横軸に時刻をとり、縦軸に速度をとったグラフであり、時間の変化による速度の推移を表すグラフである. 図3は凹凸の特徴を示している. 凹凸のうち凸の部分は移動中の部分の速度の推移であり、凹の部分は施設

内に滞在している部分の速度の推移であり、移動中のほうが施設内に滞在している部分より速いことを示している。

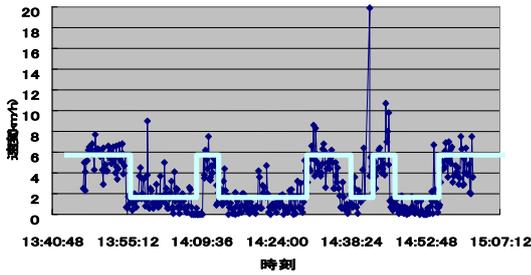


図 3 時間による速度の推移

Figure 3 Transition of the speed by time.

(2) 方向に関する知見

GPS ログ情報の隣接する 2 点間の緯度と経度の差を求めるとそれぞれ、方向を+, -, 0 の 3 パターンに分けることができる。緯度差と経度差においてそれぞれの 3 パターンを組み合わせることで表 1 のように 9 パターンの方向に分類することができる。また、表 2 の移動中の方向と施設滞在中の方向を見て分かるように、移動中の進行方向は安定、施設滞在中の進行方向は不安定ということが分かった。

表 1 方向分類表

Table 1 Direction classification table.

9パターンの方向分類		
パターン	緯度差	経度差
A	+	+
B	+	-
C	+	0
D	-	-
E	-	+
F	-	0
G	0	0
H	0	+
I	0	-

表 2 移動中と施設滞在中の方向比較

Table 2 Comparison of direction.

移動中の緯度・経度差			施設内に滞在中の緯度・経度差		
北緯差	東経差	方向	北緯差	東経差	方向
0.00005720	-0.00013730	B	-0.00004580	0.00000000	F
0.00005340	-0.00013730	B	0.00002290	0.00000000	A
0.00004200	-0.00016790	B	0.00003050	0.00003060	A
0.00005340	-0.00013730	B	0.00002670	-0.00003060	B
0.00004950	-0.00015260	B	0.00005340	0.00007630	A
0.00001910	-0.00015260	B	0.00006100	0.00000000	C
0.00004580	-0.00015260	B	0.00004580	-0.00007630	B
0.00003060	-0.00013730	B	-0.00001530	-0.00001520	D
0.00002290	-0.00010680	B	-0.00011060	-0.00003060	D
0.00004580	-0.00006110	B	-0.00018320	0.00016790	E
0.00006110	-0.00009160	B	0.00003430	0.00000000	A

5.2 有意情報の抽出アルゴリズム

(1) 抽出する有意情報

GPS ログ情報から抽出することができる情報として大別して以下の 3 つに分類することができる。

- 時系列情報 (年月日, 曜日, 時刻)
- 滞在施設情報 (施設位置・名, 滞在時間・回数等)
- 遷移情報 (移手段・経路・時間・頻度等)

滞在施設情報と遷移情報を抽出するためには、GPS ログ

情報を移動中区分と施設滞在中区分に区別する必要がある。滞在施設情報は施設滞在中区分を分析することで抽出し、遷移情報は移動中区分を分析することで抽出できる。本稿では、滞在施設情報の抽出アルゴリズムについて述べる。

(2) 移動・停止 (施設滞在) 判定アルゴリズム

5.1 節で述べた知見を基に移動・停止 (施設滞在) 判定アルゴリズムを提案する。図 4 に提案したアルゴリズムを示す。

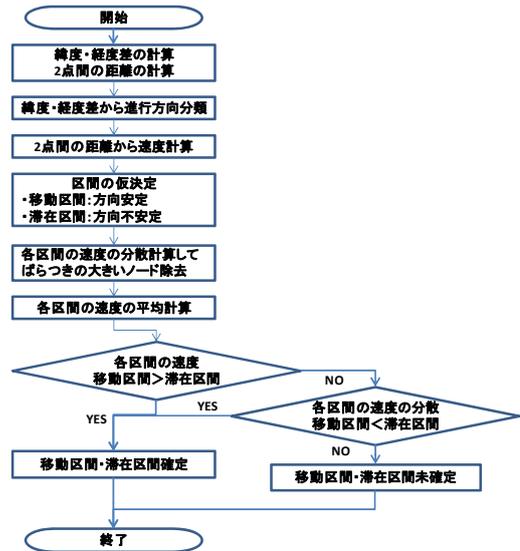


図 4 移動・停止(施設滞在)判定アルゴリズム

Figure 4 Algorithm for determining moving and standing-still sections.

(3) 緯度・経度特定アルゴリズム

(2) で区別された施設滞在中区分には複数のノードがある。その中で最もユーザの滞在施設の位置情報として最適なノードを選択する手法を 3 つ提案する。誤差の大きいノードを除去した上で平均をとる手法 (提案 a)。DBSCAN[7]適応後、施設へ入退出する瞬間のノードを含むクラスタのノードの平均をとる手法 (提案 b), 誤差の大きいノードの含有数が最小のクラスタのノードの平均をとる手法 (提案 c) を提案する。以下に各手法のアルゴリズムを示す。

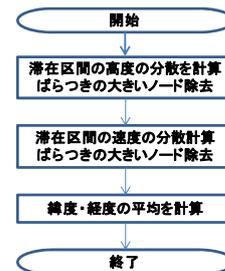


図 5 提案 a

Figure 5 Proposed method a.

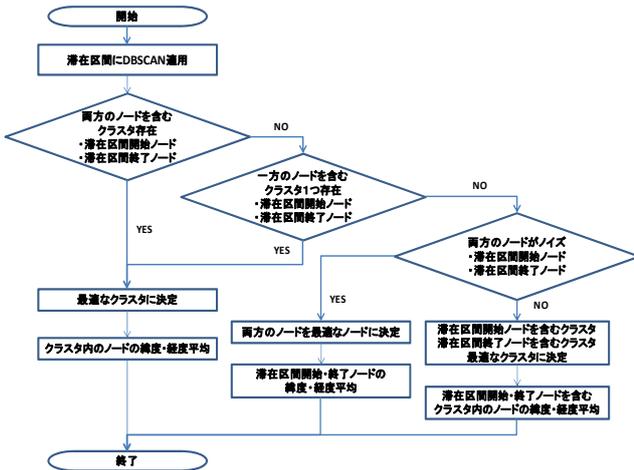


図 6 提案 b

Figure 6 Proposed method b.

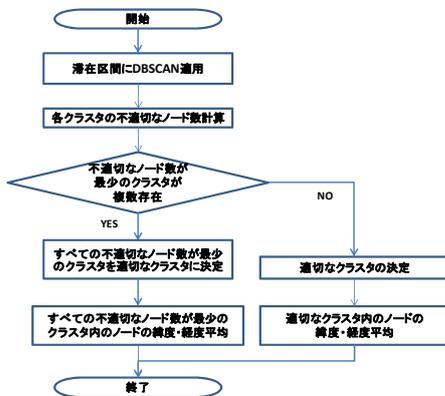


図 7 提案 c

Figure 7 Proposed method c.

5.3 習慣モデルの定義と習慣モデル導出アルゴリズム

(1) 習慣モデルの定義

人の行動は場所の遷移で表せる。そこで、人が滞在したことのある場所とその場所の遷移関係の情報を活用し、図 8 の習慣モデルを定義した。



図 8 習慣モデルの定義

Figure 8 Definition of habit model.

本研究における習慣モデルは図 8 のように習慣モデルを実現する上層と加工したデータを格納する下層による階層型構造である。ユーザの滞在した場所に関する情報を State の要素とし、その移動過程に関する情報を Link の要素とし

た。また、遷移先を決定するための要素として Tree の要素を定義し、全加工データの要素を格納するものを Log と定義した。通常は上層の情報を活用してサービスを提供するが、必要に応じて下層も利用する。

(2) 習慣モデル導出アルゴリズム概要

習慣モデルを導出するアルゴリズムを提案する。まず、LogDB の構築を行う。その後、StateDB、LinkDB、TreeDB を構築する。これらの構築手順は、登録するログと DB 上のデータと比較し、同一データがあれば頻度を更新する。もし同一データがなければ新たに DB に登録する。滞在時間等は LogDB へ問い合わせ、その結果を DB へ登録する。以下に StateDB、LinkDB、TreeDB の構築手順を図 9、図 10、図 11 に示す。

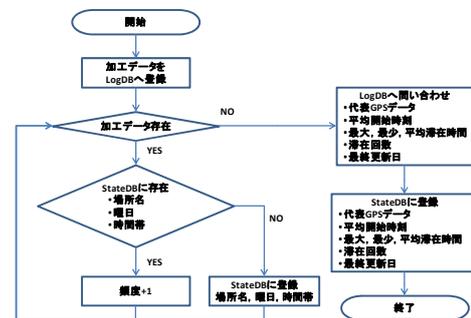


図 9 StateDB の構築手順

Figure 9 Building step of State-DB

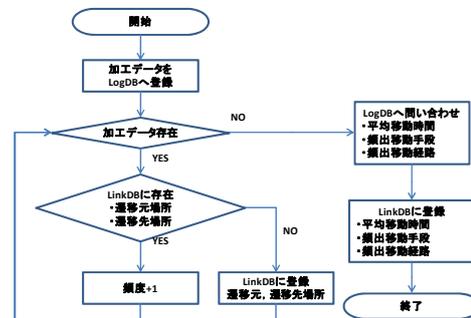


図 10 LinkDB の構築手順

Figure 10 Building step of Link-DB

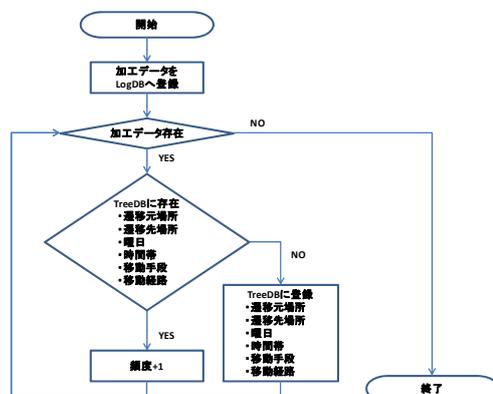


図 11 TreeDB の構築手順

Figure 11 Building step of Tree-DB

(3) 習慣モデルの遷移先決定機能

習慣モデルは場所間の遷移モデルであり、状態遷移モデルと親和性が高く、ユーザの現在の状態を契機に遷移先を決定できる。本研究では事実に基づいて遷移条件を整理した TreeDB により遷移先決定機能を実現した。曜日、時間帯、移動手段、移動経路を基にユーザが次に訪れる場所を決定する。TreeDB の要素は図 8 の Tree の要素に示している。図 8 中の TreeDB の要素において、遷移先場所が予測結果であり、曜日、時間帯、移動手段、移動経路名が遷移条件である。

5.4 習慣モデルを活用したサービス提供法

習慣モデルを活用したサービス例として、終電、終バス通知サービスを示す。以下はサービス提供の流れである。
 Step1 : StateDB の検索 (現在の滞在施設情報の取得)
 Step2 : TreeDB の検索 (目的地の決定)
 Step3 : LinkDB の検索 (移動手段の取得)
 Step4 : 現在地と自宅の最寄り駅、最寄りバス停の取得
 Step5 : Step1~Step4 で取得した情報を基に終電情報の取得
 Step6 : Step1 と同じ場所にいる場合に終電情報を通知

6. 評価

6.1 有意情報抽出アルゴリズムの評価

(1) 移動・停止(施設滞在)判定アルゴリズムの評価

移動・停止 (施設滞在) 判定アルゴリズムの判定精度に関する評価を表 3 に示す。八王子市において 25 カ所の施設へ訪れた際に収集したログ情報に対してログ情報分析プログラムへ入力し、移動・施設滞在判定に関する評価を行った。表 3 で示したように、25 カ所中 24 カ所 (96%) において施設滞在中と判定することができた。25 カ所中 1 カ所 (4%) については施設滞在中と判定することができなかった。施設滞在中と判定することができなかった場合としては、移動・停止 (施設滞在) 判定アルゴリズムにおいて進行方向を用いてログ情報を進行方向の安定部と進行方向の不安定部に正常に区別できなかった場合である。また、移動中と判定できたのは 39 区間中 38 区間 (97.4%) である。39 区間中 1 区間 (2.6%) については移動中区間と判定することができなかった。

表 3 移動・停止 (施設滞在) 判定アルゴリズム評価
 Table 3 Evaluation of algorithm for determining moving and standing-still sections.

移動・停止(施設滞在)判定アルゴリズム評価		
評価対象	移動中滞在中区間判定率	判定区間/区間数
滞在中判定	96.0%	24/25
移動中判定	97.4%	38/39

(2) 緯度・経度特定アルゴリズムの評価

移動・停止 (施設滞在) 判定アルゴリズムによって滞在中区間と判定された 24 カ所のうち Google maps が更新され

ていなかった 2 カ所を除いた、22 カ所について評価を行った。表 4 に提案した 3 つの緯度・経度特定アルゴリズムの位置特定精度の評価を示す。

表 4 緯度・経度特定アルゴリズム評価

Table 4 Evaluation of method of identifying a standing-still node from latitude and longitude information.

評価対象	緯度・経度特定アルゴリズム評価					
	提案a		提案b		提案c	
評価項目	特定率	特定施設/施設数	特定率	特定施設/施設数	特定率	特定施設/施設数
施設内特定率	31.8%	7/22	27.0%	6/22	31.8%	7/22
誤差10m以内	72.7%	16/22	63.6%	14/22	63.6%	14/22

表 4 で示すように、提案 a によって特定したノードの位置は 22 カ所中 7 カ所において施設の領域上であり、特定率としては 31.8%であった。また、誤差を 10m 考慮した場合は、22 カ所中 16 カ所あり、特定率としては 72.7%であった。提案 b によって特定したノードの位置は 22 カ所中 6 カ所において施設の領域上であり、特定率としては 27%であった。また、誤差を 10m 考慮した場合は、22 カ所中 14 カ所であり、特定率としては 63.6%であった。提案 c によって特定したノードの位置は 22 カ所中 7 カ所において施設の領域上であり、特定率としては 31.8%であった。また、誤差を 10m 考慮した場合は、22 カ所中 14 カ所であり、特定率としては 63.6%であった。以上のように満足する結果を得ることができなかった。しかしながら、提案 b と提案 c に関しては、クラスタリングは正常に行われていることを確認している。以下にクラスタリングの結果をマッピングした図を図 12 に示す。図 12 の (a) はクラスタリング前の状態であり、(b) はクラスタリング適応して、生成された各クラスタの平均位置を示している。



(a)クラスタリング前のGPSデータ (b)クラスタリング後のGPSデータ

図 12 クラスタリングの結果
 Figure 12 Results of clustering.

図 12 が示すように、クラスタリングは正常に行われている。適切なクラスタを選択することができれば、大幅に特定精度の向上が見込める。また、現段階でも提案した 3 つの手法を適切に使い分けことができれば、施設内特定率で 50%、誤差を 10m 考慮した場合で 80%程度まで精度を向上させることができる。クラスタの選択方法と各手法の使い分けができれば、本提案も十分有効性を示すことができる。

6.2 習慣モデルの有効性の評価

(1) 習慣モデルの有効性評価

習慣モデルの有効性を評価するために、既存のサービスと比較した。以下に対象の既存サービス、その既存サービスの機能を表 5 に、既存サービスに対して習慣モデルを利用した時の比較評価を表 6 に示す。

表 5 既存サービス機能

Table 5 Function of existing services

既存のサービス名	サービスの機能
コンシェルのおトGPS機能(終電アラーム)	到着駅を最大3つ登録し、現在地の最寄り駅から到着駅までの終電情報を通知し、現在地から徒歩で間に合う時間帯に配信
コンシェルのおトGPS機能(バス運行状況)	ユーザが通勤や通学に利用するバスを事前に登録し、バス停にバスが到着する時刻や最終運行のバスの時刻を配信

表 6 既存サービスとの比較

Table 6 Comparison of existing services

	コンシェル(終電)	コンシェル(終バス)	習慣モデル
初期設定	必要	必要	なし
サービス提供範囲	現在地-自宅間	登録したバス経路	滞在地-(自宅or実家)間
サービス連携	終電単独	終バス単独	終電、終バス連携可能

表 6 の習慣モデルの優位性を確認するために、習慣モデルを活用した終電通知サービスの一部を試作した。習慣モデルは移動手段を考慮するので、終電、終バスサービスを適切に、連携して提供することが可能になった。

(2) 習慣モデルのデータ量削減評価

ライフログにおいてログのデータ量の削減は重要な要素である。生の GPS ログデータから加工した情報を習慣モデルで管理することでログデータ量の削減も期待できる。生の GPS ログデータと習慣モデルを構成する各データベースのデータ量の比較を図 13 に示す。

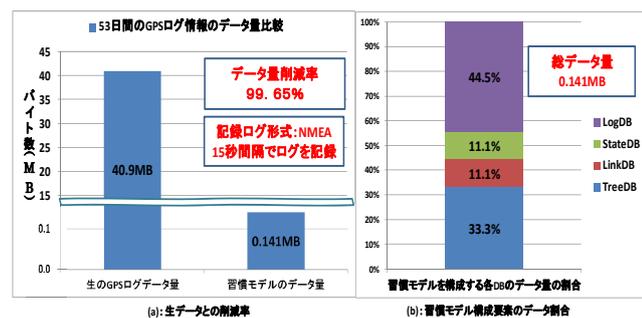


図 13 データ量削減の評価

Figure 13 Evaluation of data volume reduction.

3月中旬～4月上旬と9月上旬～11月中旬の期間に GPS Logger(tripmate850)を使用して 53 日間の間 15 秒間隔の頻度で GPS データを収集した。収集したデータの形式は NMEA(National Marine Electronics Association)である。生の GPS データと習慣モデルを構成する DB(LogDB, StateDB, LinkDB, TreeDB) の総データ量を比較した。図 13 の (a) で示すように、生の GPS データを加工することで総データ

量 40.9MB の生の GPS データを習慣モデルで管理することで、0.141MB にまでデータ量を削減することができた。割合に換算すると約 99.65%削減することができた。図 13 の

(b) は習慣モデルを構成する各 DB のデータ量の割合である。最もデータ量の割合の大きい DB は LogDB であった。LogDB は加工されたデータをすべて蓄積する DB であるので、今後 GPS ログ情報を継続的に収集していくことで今後もデータ量が増加していく傾向にあると考えられる。その他の DB である StateDB, LinkDB, TreeDB は同一の情報は頻度を更新するのみである。また、時間を経るにつれて新しい情報は少なくなる傾向にあるので、LogDB とその他の DB である StateDB, LinkDB, TreeDB とのデータ量の差がより一層大きくなると予測される。本研究では、GPS ログデータの管理方法として習慣モデルを活用することで、大幅なデータ量の削減を実現することができた。データ量の削減という観点からも習慣モデルの有効性を示すことができた。

7. おわりに

本稿では、GPS 情報から有意な情報を抽出し、有意な情報から習慣モデルを構築し、習慣モデルを活用した「気づき」サービス構成法を提案し、サービスを試作した。そして、既存サービスと比較し、習慣モデルの有効性を確認した。さらに、習慣モデルのデータ管理により大幅なデータ量の削減を実現した。

今後は、各種有意情報抽出アルゴリズムの改良、個人だけでなく複数人の習慣モデルを活用したサービス提供、GPS 以外のライフログとの連携を考慮する必要がある。

参考文献

- 岡田和也,横山輝明,門林雄基,山口英,“移動軌跡からの地点間関連性の自動推定とその応用に関する研究,”1120-1126pp,マルチメディア,分散,強調とモバイル(DICOM2009)シンポジウム 2009.7.
- 山田直治,磯田佳徳,南正輝,森川博之,“GPS 搭載携帯電話を用いた移動経路履歴に基づく訪問地予測方式,”B-15-23 電子情報通信学会総合大会 2010 年 3 月.
- 上坂 大輔,岩本 健嗣,横山 浩之,“携帯電話の位置履歴に基づく自宅職場推定法,”B-20-29 電子情報通信学会総合大会 2010 年 3 月.
- 茂木学,手塚博久,中村幸博,永徳真一郎,瀬古俊一,西野正彬,武藤伸洋,阿部匡伸,“GPS 情報を基本としたライフログ活用システムのフィールド実験,”D-9-12 電子情報通信学会総合大会 2010 年 3 月.
- Ashbrook D and Starner T, “Using GPS to learn significant locations and predict movement across multiple users”, Personal and Ubiquitous Computing, Vol.7, NO.5, pp.275-286(2003).
- 西野正彬, 瀬古俊一, 青木政勝, 山田智広, 武藤伸洋, 阿部匡伸, “滞在地遷移情報からの行動パターン抽出方式の検討”, 情報処理学会研究報告. UBI, [ユビキタスコンピューティングシステム] 2008(110), 57-64, 2008-11-06.
- Martin Ester, Hans-Peter Kriegel, Jörg Sander, Xiaowei Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise," KDD'96,1996.